Fake News in Spanish: Towards the building of a Corpus based on Twitter

[1]Braulio Andres Soncco Pimentel, [2]Roxana L.Q. Portugal

[1]UNSAAC, Cusco, Perú
[2]PUC-Rio, Rio de Janeiro, Brasil
992182@unsaac.edu.pe, rportugal@inf.puc.rio.br

Abstract. Nowadays, we can use various techniques to detect fake news that depends mainly on supervised models that trained a corpus of true and fake news in English. However, these models are not suited to classify news in Spanish. On the other hand, very few is known about the provenance of news in existing corpora. This work presents a strategy to create a Corpus of news in Spanish for fake new detection purposes. We propose Twitter as a mediator to find relevant sources of where to take the news. As such, we report the work in progress performed to characterize Twitter as a means to retrieve the news.

Keywords: Corpus, Fake News, Politics, Social Networks.

1 Introduction

Fake news is a common term in these days, however even we are conscious of its existence, we, as humans, are not as accurate to detect lies [1]. Shu et al. [2] cites that human is vulnerable to identify fake news due two factors: Naive Realism and Confirmation Bias. The first when readers believe as accurate information the ones aligned with their point of view, and the second when readers are abler to receive content that confirms their view.

Several efforts emerged to tackle the automated identification of fake news in different areas, particularly in politics [3]-[8]. However, such solutions are not as effective for classifying true and fake news in languages other than English. That is, the language characteristics influence in automated classifiers [9]. We verified few work creating Spanish corpora [10][11]. One of them [11] corroborates [9] stressing the importance of a corpus creation as a first step to deal with fake news. On the challenges to building a corpus, Tschiatschek et al. [12] indicate the lack of corpora pointing the provenance of news, as well as the subjectivity in annotations of news.

Fake news proliferates rapidly on social networks [4][9][12], as readers share the news without paying attention if they are fake [2][12]. Less attention is paid to the sources generating fake news [12]. A principle towards the building of a corpus [13] indicates that the Universe of Discourse (contents) should be properly selected to support the purpose. However, existing corpora is created by crawling news from the most established newspapers [4][9] leaving out other digital sources such as new digital newspapers and blogs.

A strategy to find relevant sources of fake news may be to identify the most widespread news on social networks such as Twitter. That is, by finding the comments that impact more on a topic, it is possible to trace the comments with URLs. Although there is a growth in the detection of fake news using the content of Twitter itself [14][15][16], this work proposes its use only as a mediator due to the nature of the tweets, which are short, fragmented and decontextualized [17].

This work presents the work in progress towards the building of a Spanish corpus for fake news identification. The paper is organized as follows. In Section 2, we explain the research goals. Section 3 details the progress to characterize Twitter in the strategy. Section 4 concludes and summarizes future work.

2 Research Objectives

To define our strategy, we elicit fake-news works [3]-[12] to know more about the definitions that we summarized in Table 1. We looked for the pros and the perils of Artificial Intelligence techniques, as such some authors stress the importance of explainable data when using AI techniques [18][19], that is, we should not rely on predictive models that do not show us their rationale.

Table 1. Eliciting Concepts about Fake News

Technical Concepts	Topics
Linguistic features, Machine learning, Veracity assessment, news verification, SVM, Knowledge networks, Predictive modelling, Feature extraction, Linguistics correlation, Visualization, Metadata	Politics, Sports, Technology, Show Business

To delineate a strategy to find relevant sources in Twitter messages, we found necessary the understanding of the paradigm of this environment. However, few details are given. The main focus is the influence of Twitter in the dissemination of news. By using Google Scholar, we searched the query *allintitle: twitter paradigm*, with which we found eight papers and selected four by its title [21]-[24]. We elicit the frequent topics: politics, sports and nursing. An important fact is mentioned, that is, a single person can easily reach thousands of contacts through a topic of interest. However, while Twitter metadata such as *tweets*, *retweets*, *hashtags*, among others are implicit, works [21]-[24] lack details about the interactions among the actors, as well as details on how a topic becomes relevant (a trend).

Following, we developed two hypotheses to tackle our main goal, which is the identification of fake news in Spanish. 1) A corpus of news in Spanish needs to be reliable, and 2) Twitter can be useful as a mediator to find of news. This hypothesis answers two questions:

What would be a reliable Corpus of news?

Any fake-news work has to build or rely on a dataset based on a corpus. A corpus creation takes time because it is needed to ensure the quality of information when collecting and classifying data. In this regard, few works [4][9][20][21] report on process and data provenance to build a classified corpus. Some authors [18][19] stress

that AI approaches need to be explainable from the inputs to outputs. Such is that an AI system is classified as opaque, interpretable, and comprehensible [19]. A corpus comes to be the input of an AI system, and, by exposing their contents' sources and the process to select them and annotate, the corpus can be less opaque.

Another aspect to consider when building a corpus is a principle stated by Sinclair & Wynne [13]: "The contents of a corpus should be selected regardless of their language, but according to their communicative function in the community in which they occur". Thus, contents of a Corpus shouldn't be selected arbitrarily but pointing a communicative function be it a topic (e.g. politics) or a type of information (e.g. news).

Therefore, we consider a reliable corpus to be one that makes its construction process transparent, as much as it characterizes a Universe of Information (UdI).

Why Twitter is relevant as a means of information?

Twitter is the ultimate way to get real-time information about various events, from politics, health, sports or natural phenomena. In the particular case of politics, authorities transmit their decisions, opinions and communicate to the population about their management. The journalists also show not only their journalistic line, but also their personal opinion [8]. Messages on Twitter, called *tweets,* are concise and easy to read and interpret. There is a notion of *hashtags*, which serves to broaden and spread all messages related to a certain topic. Such spread, when carried out by relevant political actors, can influence the people's decision [15]. There is also the citizen side, who are information consumers, who decide to accept, deny or ask questions about the *tweets* [7]. Twitter is used to generate political polarization [25], a situation that can be envisaged [26][27].

Some particular scenarios are the political choices of countries; political campaigns are transferred to social networks, reaching connected users. The best-known cases are the 2008 presidential election in the United States [28], and the influence of Facebook in the US presidential election [14].

Research Strategy

We defined seven steps towards the detection of fake news in Spanish:

1. Understand Twitter paradigm
2. Select a UdI to extract trends from Twitter
3. Identify UdI trends in the last 3 years
4. Identify at least ten relevant media in the last 3 years
5. Retrieve news from relevant media.
6. Classify true and false news using literature techniques.
7. Verify results against human review.

3 Work in Progress

So far, we have reach three steps from our strategy, despite technical constraints of Twitter. To identify the relevant tweets that may anchor the URLs of news, we modeled the Twitter actors involved in spreading the news. A script to collect the trends (*hashtags*) is made available.

3.1 Twitter Paradigm

As our focus is to investigate the spreading of fake-news on Twitter, so that we design a strategy to find relevant tweets, we depart by identifying the actors and interaction involved. By using the intentional modeling language i* [29] (Fig.1) we were able to show the different roles that Twitter user can perform, that is, in the fake-news context, we identified four social actors: politicians, journalists, citizen and anonymous. Each one can cover the place of a software agent called Twitter user. Any Twitter user can become an influencer or a troll, or both at the same time. A troll is an internet user who intentionally provokes other users, often using offensive language to attack an argument or a person. Trolls can work alone or together. A group of trolls can easily spread fake news [30]. An Influencer is a person who possesses a set of personal and social attributes so that he can reach others quickly [31]. According to [7], any Twitter user, as information consumers, decide to occupy a position: accept, deny or ask questions about the *tweets*.

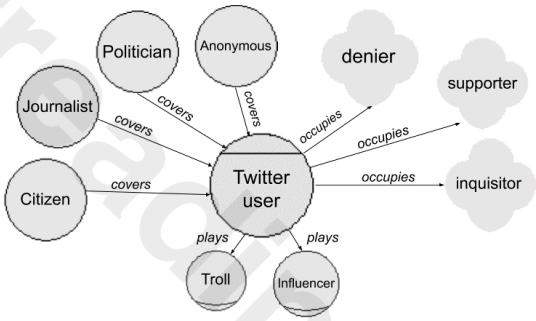

Figure 1. Actors in Twitter.

3.2 Recovering Twitter Trends

One way to filter the relevant *tweets* may be the identification of trends so we can anchor the tweets with more impact and through them to know the sources (URLs), as well as actors spreading the news. However, the Twitter API does not provide a method to get the *hashtags* that occurred in a set of time. As such, by using web-scraping techniques (https://github.com/soncco/hashtag-scraper) we collect them from Trendogate.com. This site keeps a trend archive of different locations since March 2015. Thus, we can recover a set of *hashtags* given a UdI: [Location, Time interval].

As a sample, in Fig. 2, we configured the UdI: Lima-Peru from March 2018 to March 2019. After recovering, we ordered the hashtags by frequency and selected the top 100. From them, we manually selected the ones related to the political topic.

'César Villanueva', 'Alberto Fujimori', 'Daniel Salaverry', 'Pedro Chávarry', '#NiUnaMenos', 'Edwin Oviedo', '#Elecciones2018', 'Luz Salgado', '#GlatzerTuesta', '#LavaJato', 'Rosa Bartra', 'Kenji Fujimori', '#Congreso', 'Presidente Vizcarra', 'Fuerza Popular', 'Alejandro Toledo', 'Fiscalía', 'Yesenia Ponce', 'Mesa Directiva'

Figure 2. Hashtags about politics in in Lima-Perú, from March 2018 to March 2019

Twitter Constraints

The Twitter API provides a way to get *tweets* according to premium[1] parameters and operators for accessing data. One can get *tweets* from the beginning of Twitter (2006) to the present by combining them. However, a free developer account only allows 50 queries per month, as well as limiting the use of some operators.

4 Conclusion

This work presents the steps achieved to build a reliable corpus of news in Spanish for fake-news detection. We describe the strategy to cope with Twitter data as a mediator to find sources of news. As *hashtags* are a means to find relevant tweets, we have worked in an automated way to retrieve them.

Future work is focused on strategies to semi-automate the *hashtags* filtering given a topic (Fig. 2), as well as modeling each actor intentions in the spreading of news (Fig. 1). By modeling the Twitter paradigm in the fake-news context, we will be able to create heuristics to identify relevant sources. In parallel, given the restrictions of the Twitter API, we will work on mechanisms to create efficient queries with the *hashtags* that will lead us to retrieve relevant tweets.

Acknowledgment

Portugal acknowledges the support of Cnpq scholarship 146372/2018-2.

References

1. Bond Jr, C. F., & DePaulo, B. M.. Accuracy of deception judgments. Personality and social psychology Review, 10(3), 214-234. (2006).
2. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36. (2017).
3. Karimi, H., Roy, P., Saba-Sadiya, S., & Tang, J. Multi-source multi-class fake news detection. In Proceedings of the 27th International Conference on Computational Linguistics (pp. 1546-1557). (2018, August).
4. Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. Automatic detection of fake news. arXiv preprint arXiv:1708.07104. (2017).
5. Shu, K., Wang, S., & Liu, H. Understanding user profiles on social media for fake news detection. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR) (pp. 430-435). IEEE (2018, April).
6. Choraś, M., Giełczyk, A., Demestichas, K., Puchalski, D., & Kozik, R. Pattern Recognition Solutions for Fake News Detection. In IFIP International Conference on Computer Information Systems and Industrial Management (pp. 130-139). Springer, Cham (2018, September).

[1] https://developer.twitter.com/en/docs/tweets/search/guides/standard-operators

7. Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., & Procter, R. Detection and resolution of rumours in social media: A survey. ACM Computing Surveys (CSUR), 51(2), 32 (2018).
8. Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. Predicting elections with twitter: What 140 characters reveal about political sentiment. In Fourth international AAAI conference on weblogs and social media. (2010, May).
9. Monteiro, R. A., Santos, R. L., Pardo, T. A., de Almeida, T. A., Ruiz, E. E., & Vale, O. A. Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. In International Conference on Computational Processing of the Portuguese Language (pp. 324-334). Springer, Cham. (2018, September).
10. Posadas-Durán, J. P., Gómez-Adorno, H., Sidorov, G., & Escobar, J. J. M. Detection of fake news in a new corpus for the Spanish language. Journal of Intelligent & Fuzzy Systems, 36(5), 4869-4876. (2019).
11. Blázquez-Ochando, M. El problema de las noticias falsas: detección y contramedidas. (2018).
12. Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., & Krause, A. Fake news detection in social networks via crowd signals. In Companion of the The Web Conference 2018 on The Web Conference 2018 (pp. 517-524). International World Wide Web Conferences Steering Committee. (2018, April).
13. Sinclair, J., & Wynne, M. Corpus and text-basic principles' in developing linguistic corpora: a guide to good practice. (2004).
14. Allcott, H., & Gentzkow, M. Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), 211-36. (2017).
15. Bastos, M. T., Raimundo, R. L. G., & Travitzki, R. Gatekeeping Twitter: message diffusion in political hashtags. Media, Culture & Society, 35(2), 260-270. (2013).
16. Buntain, C., & Golbeck, J. Automatically identifying fake news in popular twitter threads. In 2017 IEEE International Conference on Smart Cloud (SmartCloud) (pp. 208-215). IEEE. (2017, November).
17. Gounari, P. Authoritarianism, Discourse and Social Media: Trump as the 'American Agitator'. Critical Theory and Authoritarian Populism, 207. (2018).
18. Doran, D., Schulz, S., & Besold, T. R. What does explainable AI really mean? A new conceptualization of perspectives. arXiv preprint arXiv:1710.00794. (2017).
19. Ghosh, P. AAAS: Machine learning "causing science crisis." Retrieved June 12, 2019, from https://www.bbc.com/news/science-environment-47267081. (2019, February 16).
20. Portugal, R.L.Q., Roque, H. and do Prado Leite, J.C.S., September. A Corpus Builder: Retrieving Raw Data from GitHub for Knowledge Reuse In Requirements Elicitation. In SIMBig (pp. 48-54) (2016).
21. Moulai, H., & Drias, H. Towards the Paradigm of Information Warehousing: Application to Twitter. In International Conference on Computer Science and its Applications (pp. 147-157). Springer, Cham. (2018, April).

22. Basaille, I., Kirgizov, S., Leclercq, É., Savonnet, M., & Cullot, N. Towards a Twitter observatory: a multi-paradigm framework for collecting, storing and analysing tweets. In 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS) (pp. 1-10). IEEE. (2016, June).
23. Sheffer, M. L., & Schultz, B. Paradigm shift or passing fad? Twitter and sports journalism. International Journal of Sport Communication, 3(4), 472-484. (2010).
24. Lesen, A. E. a new paradigm For Science communication? Social media, twitter, Science, and public engagement: A Literature Review. Scientists, Experts, and Civic Engagement: Walking a Fine Line, 111. (2015).
25. Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. Political polarization on twitter. In Fifth international AAAI conference on weblogs and social media. (2011, July).
26. Colleoni, E., Rozza, A., & Arvidsson, A. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. Journal of communication, 64(2), 317-332. (2014).
27. Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. Predicting the political alignment of twitter users. In 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing (pp. 192-199). IEEE. (2011, October).
28. Cogburn, D. L., & Espinoza-Vasquez, F. K. From networked nominee to networked nation: Examining the impact of Web 2.0 and social media on political participation and civic engagement in the 2008 Obama campaign. Journal of Political Marketing, 10(1-2), 189-213. (2011).
29. Yu, E. Strategic modelling for enterprise integration. In Proceedings of the 14th world congress of the international federation of automatic control (pp. 5-9). (1999, July).
30. Dalton, M. D., & Luke, J. S. U.S. Patent No. 9,961,115. Washington, DC: U.S. Patent and Trademark Office. (2018).
31. Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. Everyone's an influencer: quantifying influence on twitter. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. 65-74). ACM. (2011, February).